KEKER
VAN NEST
&PETERS   LATHAM&WATKINS LLP   IIIORRISON FOERSTER

March 31, 2025

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl St.
New York, NY 10007-1312

cc: All Counsel of Record (via ECF)

Re: ***Discovery dispute concerning The Times's date range for custodial collection***

Dear Judge Wang:

OpenAI requests a pre-motion conference to address The New York Times's refusal to expand
the date range for its Google Workspaces and Slack collection for six key custodians: (1) A.G.
Sulzberger, Chairman and Publisher; (2) Meredith Levien, Chief Executive Officer; (3) William
Bardeen, Executive Vice President and Chief Financial Officer; (4) Greg Miller, Executive
Director Business Development and Operations, nytLicensing; (5) Alex Hardiman, Chief
Product Officer; and (6) David Perpich, Publisher of the Athletic.

The Times initially took the position that only its communications post-dating the release of
ChatGPT were relevant and accordingly applied a narrow date range limitation of November
2022 to May 2024—18 months—to its entire custodial collection. The Times did so even while
making sweeping claims regarding the existential threat Defendants' LLM products pose to its
over 170-year-old business, basing its claims on alleged misconduct pre-dating November 2022,
and seeking discovery from OpenAI from 2015 onwards. Over the course of extensive meet and
confer about this issue, The Times has budged slightly, agreeing to expand collection for four of
its 22 custodians back to 2018. Yet, it refuses to expand collection for the six custodians
identified above back to 2018, making vague allegations of burden that contradict the testimony
of The Times's Rule 30(b)(6) deponent regarding its custodial collection. The Times should be
ordered to collect documents from Google Workspaces and Slack for these six custodians back
to 2018, a request which is narrowly tailored to unearth documents relevant to OpenAI's statute
of limitations defense and The Times's allegations of harm.

I.   The dispute over date range is ripe for adjudication.

OpenAI first raised the issue of The Times's narrow custodial date ranges in the parties' October
17, 2024 joint agenda, filed prior to the October 30 discovery conference. ECF 262 at 15. The
parties agreed to meet and confer further regarding this issue after that hearing and made modest
progress: The Times expanded its custodial collection for Sam Felix, Gordon Diggs, Chris
Wiggins, and Rebecca Grossman-Cohen to 2018, its collection for Hannah Poferl to 2019, and
agreed to collect 2021 data for several other custodians. But The Times refused to collect data
back to 2018 from key c-suite custodians such as Sulzberger, Levien, and Bardeen. Instead, The
Times offered to conduct limited, "manual" searches in custodians' email in lieu of collecting
their documents and running the search terms the parties have spent months negotiating. Ex. 1
(12/23/24 Peaslee email). OpenAI raised the issue of The Times's insufficient custodial

2897306

collection again prior to the January 22, 2025 hearing. ECF 395. The Court ordered the parties to meet and confer on this and other issues (ECF 441), which OpenAI has diligently pursued, with no change in The Times's position. *See, e.g.*, Ex. 2 (2/24/25 Deamer email); Ex. 3 (3/27/25 Deamer email). Given the parties' inability to resolve this dispute over the past six months, and the advanced stage of document discovery, this dispute is ripe for Court intervention.

II.   The date range expansion is necessary for the collection of documents related to OpenAI's statute of limitations defense and The Times's allegations of harm.

The six custodians for which OpenAI seeks expanded collection are likely to have pre-2022 documents relevant to a variety of subjects, but most crucially, to OpenAI's statute of limitations defense and The Times's allegations of harm. As OpenAI alleged in its Motion to Dismiss, OpenAI disclosed, no later than July 22, 2020, that it had used a "filtered version of Common Crawl" to train its GPT-3 model, and it was "common knowledge" that Common Crawl datasets included Times articles. ECF 52 at 6. Judge Stein recently denied OpenAI's Motion to Dismiss on statute of limitations grounds, deeming it a question of fact when the Times became aware of OpenAI's alleged conduct. Ex. 4 (1/14/25 Hr'g Tr. at 36:13-22). Yet, The Times is preventing OpenAI from gathering the facts from the high-level custodians whose knowledge is most likely to be imputed to the company—its Chairman, CEO, and CFO, as well as custodians like Hardiman and Perpich, who hold key product roles.

These custodians are also likely to have responsive documents related to the impact of Defendants' LLMs on The Times's business model—indeed, they are already designated as custodians for search terms related to these subjects. The Times summarily claims, however, that there is no need to collect earlier data from these custodians to run these search terms against, because "no one at The Times was discussing or documenting [the impact of Defendants LLMs] before ChatGPT was publicly released." ECF 262 at 16. As an initial matter, The Times own reporting suggests otherwise: e.g., in July 2020, opinion columnist Farhad Manjoo published an article discussing GPT-3's writing capabilities and the potential implications for The Times's journalists. *See* Ex. 12 (Manjoo, *How Do You Know a Human Wrote This?*). And if it is true that these custodians were not having such conversations privately, there would be no additional documents for The Times to review. But, The Times cannot avoid additional collection by claiming, *a priori*, that there is nothing to see here.

III.   The Times has not substantiated its burden allegations.

Despite the parties' months of meet and confer over this issue, The Times has failed to substantiate its claims that conducting this limited additional collection would be unduly burdensome. *See Zhulinska v. Niyazov Law Grp., P.C.*, No. 21-CV-1348, 2021 WL 5281115, at *3 (E.D.N.Y Nov. 12, 2021) (a party cannot evade its discovery obligations by offering "merely speculative" concerns regarding burden). In rejecting OpenAI's compromise offer, The Times has said only that "more than a million messages are stored in Google Mail," without specifying whether those emails belong to the custodians at issue or providing hit counts reflecting the potential review burden. Ex. 5 (3/28/25 Muttalib email).

The Times has represented that its ability to collect documents, run search terms, and provide hit counts is limited. *See* Ex. 6 (1/22/25 Hr'g. Tr. at 55:8-11) ("The Times is not available or able, due to the software that it uses, to run search terms in the native environment."); Ex. 7 (12/3/24 Hr'g. Tr. at 36:22-24, 39:10-11) ("We have not provided hit counts from [OpenAI's requested] custodians . . . because in order to provide the hit counts," The Times would have to "pause the process of collecting documents from the [custodians] we have agreed to.").

However, OpenAI took The Times's Rule 30(b)(6) deposition with respect to custodial collection on January 30, 2025. ███████████████████████████

███

Ex. 8 (1/30/25 Mandel Tr. at 76:6-7, 75:25-76:2)

██████████████████████████). Google Vault allows users to "search [their] organization's Google Workspace data by user account, organizational unit, date, or keyword," and also allows for easy collection of "a comprehensive copy of all the data that matches [the] search criteria." Ex. 9 (Vault - Google Vault Help). It further allows for the collection of hit counts prior to an export, to help organizations "refine [their] query filters to return more or less results." Ex. 10 (Google Vault | Google for Developers).

Since the last discovery conference in January, OpenAI has repeatedly asked The Times to explain the discrepancy between its representations in court regarding burden and its Rule 30(b)(6) custodian's testimony. *See, e.g.* Ex. 2 (2/24/25 Deamer email). Contrary to its in-court assertions, The Times should be able to run search terms in the "native environment," given that Google Vault explicitly allows for that capability. The Times also should be able to provide hit counts—another standard Google Vault capability—without pausing all document collection. Yet, over the last two months of meet-and-confer communications, The Times has declined to clarify, or to otherwise substantiate the claims it made to the Court. And, The Times's refusal to produce additional Slack data rests on even shakier ground. The Times experienced an issue with its Slack data production leading to certain date discrepancies, an issue The Times has been able to correct with an overlay. Ex. 11 (3/21/25 Muttalib email). On the basis of that technical difficulty alone, The Times now claims any additional Slack collection is unduly burdensome.

Having failed to provide meaningful information regarding burden during meet and confer, The Times should not be able to hide behind burden to avoid the limited expansion of collection OpenAI requests. This is especially true because OpenAI has limited its request to six of The Times's 22 custodians, is only seeking Google Workspaces and Slack data for those custodians, and has not requested that The Times re-image any work or personal devices to comply with this request. Notably, The Times is the only party that has applied such a narrow date range to its custodial collection. The other News Plaintiffs have agreed to collect documents back to 2018, and The Times has sought—and obtained—discovery from OpenAI going back to its 2015 founding. Nor is there any basis for The Times's request that OpenAI develop a new set of terms for pre-2022 collection. *See* Ex. 5 (3/28/25 Muttalib email). The parties have extensively negotiated search terms, and those terms should be applied to any additional custodial collection. For the foregoing reasons, OpenAI respectfully requests that the Court order The Times to conduct the limited custodial collection requested for the six custodians enumerated above.

2897306

Respectfully,

KEKER, VAN NEST & PETERS LLP[1]	LATHAM & WATKINS LLP	MORRISON & FOERSTER LLP
/s/ Sarah Salomon	*/s/* Herman Yue	*/s/ Rose S. Lee*
Sarah Salomon	Herman Yue	Rose S. Lee

---

[1] All parties whose electronic signatures are included herein have consented to the filing of this document.

2897306